A CISCO INTERNAL WHITE PAPER

CISCO WIDE AREA APPLICATION SERVICES TECHNICAL OVERVIEW

EXECUTIVE SUMMARY

Enterprise organizations face numerous challenges to the delivery of applications and critical business data to the enterprise edge. As the global workforce continues to become more distributed, providing adequate service levels throughout the organization becomes increasingly difficult, causing I/T organizations to deploy costly and difficult-to-manage infrastructure at each location, including file services, email, video, software distribution, print services, and more. With continuous pressure from industry and federal regulation, I/T organizations now face the significant challenge of consolidating this costly infrastructure to improve data protection and compliance with government and industry regulation while somehow maintaining the service level the distributed workforce has come to expect. Furthermore, as applications continue to become more complex and increase in size, delivering applications and application information to the enterprise edge becomes an increasing challenge. Cisco Wide Area Application Services (WAAS) is a new breed of technologies that enable application delivery and infrastructure consolidation while leveraging existing capital and operational assets.

CHALLENGE

Information Technology (IT) budgets are not increasing with the growing expectations executives have on IT resources. As such, consolidating costly remote office infrastructure to contain capital and operational expenditure is highly desired by almost every IT organization that manages a distributed infrastructure. Furthermore, as applications continue to evolve and become larger and more complex, the network load created grows and the performance characteristics of the Wide Area Network (WAN) impact application delivery even more so. The challenges of data retention policies, business continuance, and disaster recovery requirements further exacerbate the problem given a heavily distributed infrastructure and already over-burdened WAN environment. Having a centralized IT infrastructure helps enable operational and capital cost savings while streamlining data protection processes.

Until this point, many vendors attempted to solve such problems with point products that did not effectively allow IT organizations to leverage existing investment in network intelligence.

CISCO WAAS OVERCOMES THE WAN

Cisco WAAS is a solution that uses a symmetric approach to application optimization, that is, a device on each side of the WAN that has application-specific intelligence and network-specific intelligence. These devices, called the Cisco Wide Area Application Engine (WAE), shown in figure 1, are available as a router-integrated network module or as standalone appliances, and are deployed out of the data-path in the data center and in the remote office LAN.

By deploying Cisco WAAS, IT organizations will find that they are better positioned to:

- Centralize costly distributed IT capital resources into the data center
- Improve throughput and delivery of applications and application data to the enterprise edge

- Increase efficiency for existing WAN connections
- Maintain remote office user application performance expectations

Cisco WAAS enables such benefits through a series of optimizations that are not only application-friendly, but also packet-network friendly.

- Robust application-specific and protocol-specific adapters that mitigate latency, suppress unnecessary messages, and provide protocol-specific data and metadata caching to minimize unnecessary data transfers and messaging over the WAN to improve interactive application responsiveness
- Advanced protocol-agnostic network compression capable of removing redundancy from TCP traffic regardless of application and provide computational compression
- Standards-based throughput improvement technologies and advanced congestion management for TCP-based applications and transmissions
- Local infrastructure services to minimize the amount of administrative traffic using the WAN

With Cisco WAAS, almost any TCP-based application can benefit from the network and application-specific acceleration techniques, including Internet/Intranet applications, databases, file services, file transfer, email, data protection, client-server applications, and many others.

The Cisco WAAS software platform runs on the high-performance and extensible Cisco Wide Area Application Engine (WAE) family of appliances or router-integrated network module (NM-CE), shown below in figure 1:

Figure 1. Cisco Wide Area Application Engine (WAE) appliances and router-integrated network module.

Cisco Wide Area Application Services (WAAS) helps to enable infrastructure consolidation while improving WAN utilization efficiency and application delivery through the following application acceleration and WAN optimization features:

- **Transport Flow Optimizations (TFO)** – TFO provides standards-based, field-proven throughput improvements for TCP-based applications while maintaining packet-network friendliness and safe coexistence with other network nodes communicating using standard TCP implementations
- **Data Redundancy Elimination (DRE)** – DRE is a bidirectional database of traffic that has been previously-seen traversing the network. DRE inspects incoming TCP traffic and identifies patterns within the message. Once

patterns have been identified, redundant patterns can be safely replaced by small signatures, thus minimizing the bandwidth consumption of each message significantly. DRE not only minimizes the amount of traffic that must traverse the WAN, but also maintains full protocol and service coherency, as the original message is fully rebuilt by the distant WAE and verified for accuracy. As DRE works within the transport layer, it is application protocol agnostic, and patterns that have been marked and indexed for one protocol can be leveraged when that same data is sent using another protocol. DRE can provide anywhere from 2:1 to 200:1 compression based on the application being used and data being transmitted.

- **Lempel-Ziv (LZ) Compression** – LZ compression is a standards-based compression that can be employed to further minimize the amount of bandwidth consumed by a TCP flow. LZ compression can be used in conjunction with DRE or independently. LZ compression can provide anywhere from 2:1 to 4:1 compression based on the application being used and data being transmitted.
- **Application Traffic Policy (ATP)** – ATP gives administrators flexibility in configuring how specific application protocols are optimized by Cisco WAAS. While Cisco WAAS ships with default policies for over 100 different traffic types and more than 25 application groups, the administrator can modify existing policies or create new policies to match other application flows found in the WAN.
- **Industry-leading Wide Area File Services (WAFS)** – Cisco WAAS builds upon the robust Wide Area File Services capabilities provided by the Cisco File Engine family. Cisco WAAS provides extensive file services capabilities for CIFS and NFS clients at the network edge, and can safely mitigate unnecessary protocol messaging and data transfer, thereby greatly improving the remote user experience. Cisco WAAS also provides Windows-compatible edge print services with centralized driver management and distribution. Furthermore, like any application adapter that can run on Cisco WAAS, the Wide Area File Services application adapter can leverage the throughput improvements and compression provided by WAAS.
- **Extensible application platform** – Cisco WAAS is designed to meet current and future application delivery and infrastructure consolidation challenges. The modular software architecture provided by Cisco WAAS allows for additional robust application-specific adapters to be seamlessly integrated.
- **Deployment Flexibility and Availability** – Cisco WAAS is the only application delivery platform today that offers deployment flexibility, availability, and packet network transparency. Cisco WAAS integrates seamlessly with the packet network using network interception technologies such as the Web-Cache Communication Protocol version 2 (WCCPv2), Policy-Based Routing (PBR), and server load balancing (SLB) platforms such as the Cisco Content Services Module (CSM) for the Catalyst 6509 multilayer switch. Cisco WAAS and the WAE family of hardware platforms can be deployed in a highly-available, load-sharing configuration, and maintain full network transparency from end-to-end. Cisco WAAS transparency helps IT organizations maintain capital and operational investment in existing network features such as Quality of Service (QoS), Network-Based Application Recognition (NBAR), NetFlow, and others.
- **Network Transparency** – Cisco WAAS provides packet network transparent optimizations and preserves original packet information, including the source and destination IP and TCP information, to allow intermediary routers and switches to continue to perform functions against optimized packets such as classification, prioritization, access-control, queuing, firewall policies, NetFlow, and routing decisions.

The following sections will provide detail about each of the advanced features of the Cisco WAAS solution.

DEPLOYMENT FLEXIBILITY

Cisco WAAS application acceleration and WAN optimization is tightly coupled with the packet-network. Through the use of either WCCPv2, PBR, or the CSM, Cisco WAAS integrates seamlessly into the packet network, requiring no changes to clients or servers. With Cisco WAAS, high availability, scalability, and transparency are provided, and security, accounting, and application-specific policies are fully maintained. Once Cisco WAAS has been integrated into the packet network, it is then capable of providing robust application acceleration and WAN optimization capabilities to help enable infrastructure consolidation and WAN efficiency.

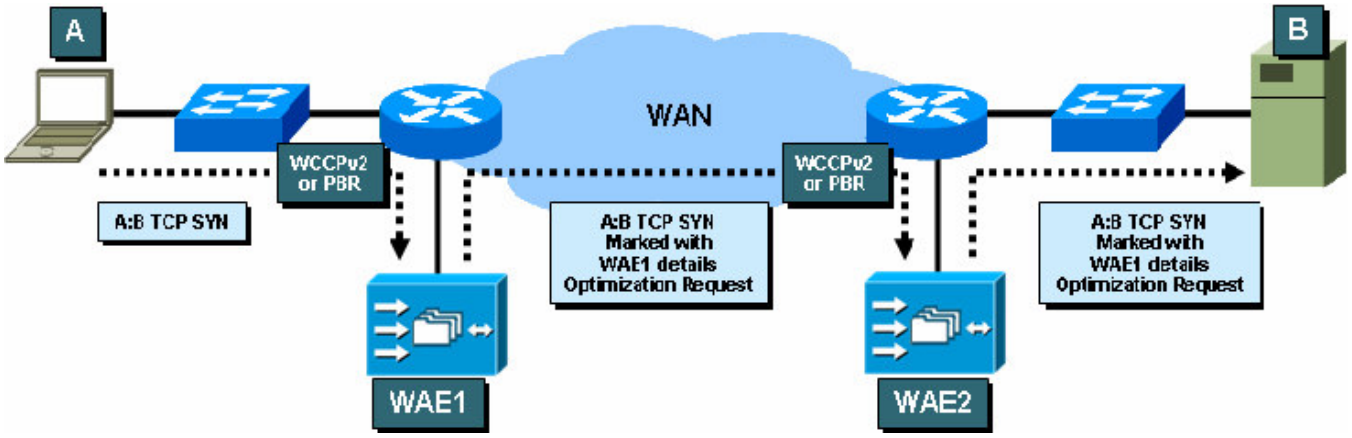
- WCCPv2, which was originally developed by Cisco, enables the transparent integration of application acceleration technology into the network with high availability and load-sharing. The Cisco WAE devices within a given location will advertise their availability to the router (or switch, multiple routers or switches can be used for network path high availability) and specify that TCP traffic should generically be forwarded to the WAE. Once the WAE devices have joined the service group with the router, the router will monitor traffic for flows that should be forwarded to the WAE instead of the original destination. As the WAE begins receiving traffic, it can then selectively apply optimizations and protocol-level handling based on application policy. With WCCPv2, up to 32 WAEs can join a service group with up to 32 routers, and each will receive a portion of the workload that would otherwise traverse the WAN unoptimized. Should a WAE fail, surviving members can take over the workload of the failed WAE. Should all WAE devices fail within a given location, traffic will be forwarded across the WAN unoptimized until a WAE is recovered.
- PBR is another deployment option available for WAAS and the WAE. With PBR, the network administrator can configure a WAE or multiple WAEs as a next-hop devices for all or specific TCP traffic. As TCP traffic is received by the router, it is forwarded to the WAE as the next-hop router, at which point it can apply optimizations based on application policy. Like WCCPv2, PBR provides transparent integration into the packet network, and also offers high-availability to the remote office or data center in that should a WAE fail, another WAE defined as a next-hop can be utilized. Should all WAEs fail, the policy-based route will be considered unavailable, and traffic will be forwarded across the WAN unoptimized until a WAE is recovered.

Once the WAE has been introduced to the packet-network, it can begin applying network and application optimizations based on application policy. For applications where an explicit application adapter is present, application-layer messaging can be terminated locally to mitigate latency and unnecessary data transfer. For any TCP traffic that must traverse the WAN, the WAE will non-intrusively mark the packets such that when received by a distant WAE (through WCCPv2 or PBR), the two WAEs can identify one another and establish peering, negotiate optimization capabilities, and ultimately, begin applying optimizations to the traffic that must traverse the WAN.

DEVICE AUTODISCOVERY

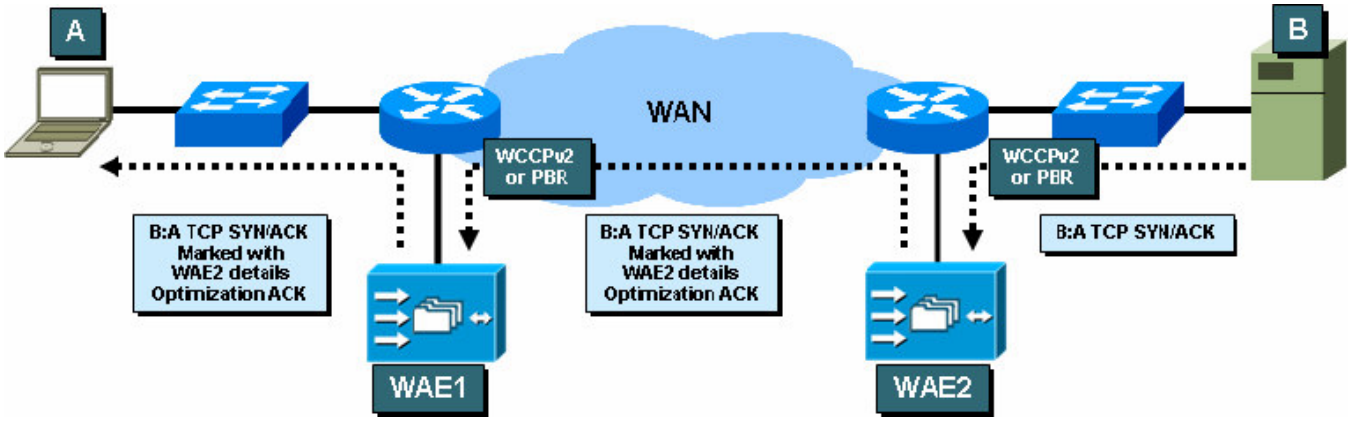
Cisco WAAS integrates transparently with the packet network and also automatically discovers all Cisco WAE devices in the path between a source and destination pair. As each TCP connection is established, Cisco WAAS non-intrusively applies markings to the connection request packets to identify each Cisco WAE in the path between the communicating nodes, as well as which optimizations are being requested based on the configured policy. As marked packets are received by distant Cisco WAEs, optimization capabilities can then be determined.

Figure 2. Cisco Wide Area Application Services Auto Discovery Process - Requestor.



When the receiving node responds to the connection request, the WAE near the receiver sees that the connection response packet is for a connection that is awaiting optimization. The WAE near the receiver then applies optimization acknowledgement markings to the connection response packets. The packets are then returned to the network for delivery to the requestor. As the packets reach the network where the requestor is located, the packets are delivered to the nearby WAE, and the WAE then understands the optimization capabilities of the other WAE in the path. At this point, both WAEs are fully aware of one another, and are able to start applying optimizations. Figure 3 shows the receiver side of the auto-discovery process.

Figure 3. Cisco Wide Area Application Services Auto Discovery Process - Receiver.



The Cisco WAE will always leave the optimization markings in the connection packets to enable support for environments where multiple Cisco WAEs may be in the path between source and destination. This allows for topologies such as full-mesh, partial-mesh, ring, star, and others. Optimization establishment always occurs between the two most distant Cisco WAE devices, even if multiple Cisco WAEs are in the path.

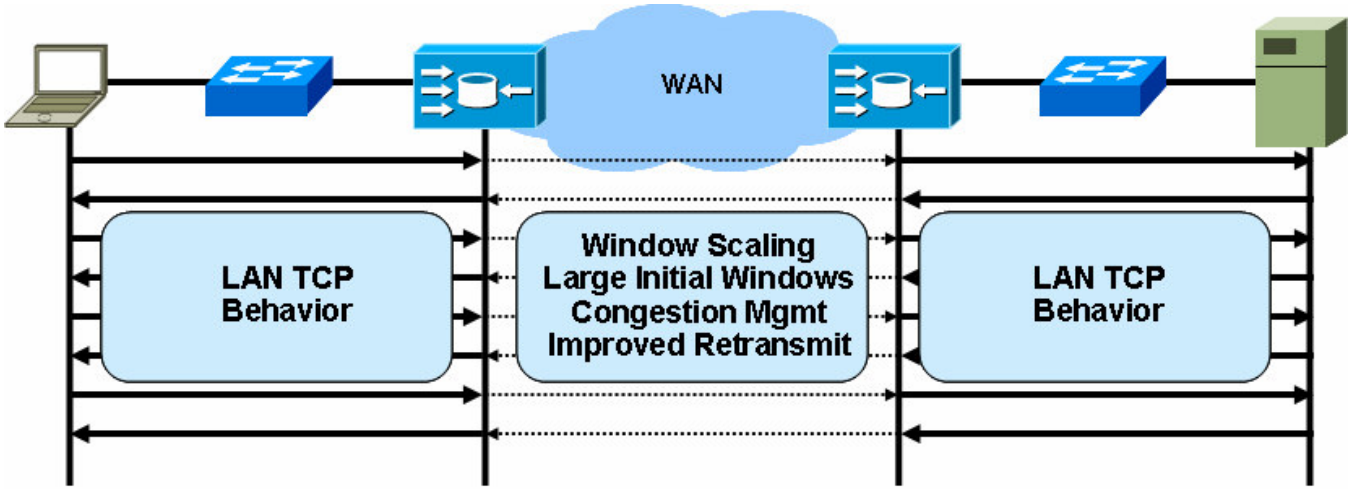
TRANSPORT FLOW OPTIMIZATIONS (TFO)

Once Cisco WAE devices have discovered one another during the client-server connection establishment, multiple optimizations can then be applied based on the configured Application Traffic Policy (ATP), which is discussed later in this document. One such optimization is Transport Flow Optimizations, or TFO, by which Cisco WAAS can help applications overcome limitations caused by the WAN and under-powered client or server TCP stacks. Cisco WAAS TFO employs a robust TCP proxy to safely optimize TCP at the WAE device by applying TCP-compliant optimizations to shield the clients and servers from poor TCP behavior due to WAN conditions. Cisco WAAS TFO improves throughput and reliability for clients and servers in WAN environments:

- Large initial windows – Cisco WAAS allows a TCP window to have an initial maximum-segment-size (MSS) of 4380 bytes, or roughly 3 times that of standard TCP. This allows short-lived connections to be completed much more quickly and efficiently by removing bandwidth starvation as a bottleneck, and provides noticeable performance improvement for all flows over high delay links.
- Window scaling – Cisco WAAS allows devices using standard TCP implementations to enjoy window-scaling capabilities to drastically improve performance over high-bandwidth, high-delay WAN links. By safely scaling TCP windows from the standard 64Kbytes to a maximum of 1GBytes, Cisco WAAS enables applications that would normally be throughput-constrained to perform well in WAN environments and fully leverage the available bandwidth provided by the WAN.
- Efficient window management – Cisco WAAS employs standards-based efficient window management technologies such as Selective Acknowledgement (SACK) to minimize the burden and performance impact of retransmission in scenarios where a packet has been lost.
- Advanced congestion management – Cisco WAAS uses advanced congestion management techniques to ensure that maximum throughput is safely restored after scenarios where packet loss has been encountered. Cisco WAAS advanced congestion management not only helps improve overall throughput, but also maintains friendliness with other TCP implementations (including standard TCP implementations) that may be in use on the network.

Figure 4 shows how Cisco WAAS TFO shields clients and servers form WAN conditions. Cisco WAAS TFO provides significant performance improvement and stability for clients and servers, as standard operating system TCP stacks were not designed to operate in WAN environments.

Figure 4. Cisco WAAS TFO Improves Application Performance and Reliability



.

DATA REDUNDANCY ELIMINATION (DRE) AND LEMPEL-ZIV (LZ) COMPRESSION

Cisco WAAS employs advanced network compression to minimize the amount of data that must be transferred per connection. Cisco WAAS advanced network compression is built from two unique compression types that are capable of working independently or in conjunction with one another.

Data Redundancy Elimination, or DRE, is an advanced form of network compression that allows Cisco WAAS to maintain a database of data that has been seen previously traversing the network and use this information to remove redundant message patterns from future or current transmissions. This enables significant levels of compression for redundant traffic patterns, and also ensures message and application coherency in that the original message is always rebuild and verified by the distant WAE. Being that DRE operates within the context of the network and is bi-directional, it is not only application-agnostic, but effective regardless of the direction of traffic flow. As such, data patterns that have been identified for one application protocol can be used by other applications, and patterns that have been identified for one direction of traffic flow can be used to remove redundancy for traffic flowing in a different direction. With DRE, a user can access information through one protocol or application and receive significant compression when accessing the same or similar information through a completely different protocol or application.

With each message received by a Cisco WAE, a signature of the entire message is computed and used for message validity verification when the encoded message is decoded by another Cisco WAE. After calculating the message validity signature, the message is then broken into content-based chunks of data. For each chunk that is identified, a signature is generated. After the entire message has been analyzed and signatures for chunks have been generated, the Cisco WAE will then add any new signatures and associated data chunks to the DRE database. For signatures and data chunks that already exist in the DRE database, the Cisco WAE is then able to replace that entire chunk of redundant, previously-seen data with the signature itself. A new encoded message is built based on the original message and includes signatures for new data chunks to be added to the distant Cisco WAE DRE database, signatures for redundant traffic that map to existing entries in the distant Cisco WAE DRE database, and the message validity signature.

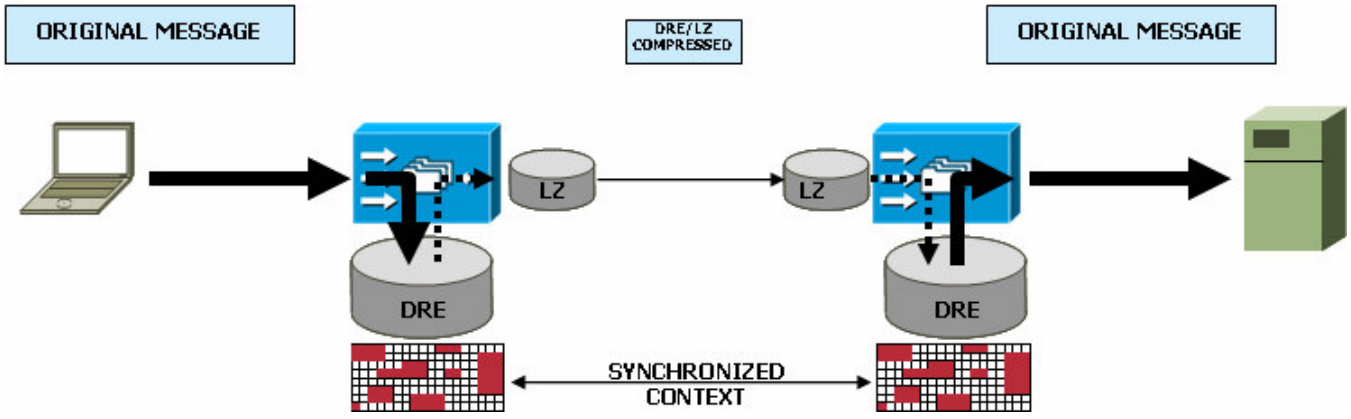
The DRE message can then be further compressed through LZ compression, if defined by the application traffic policy, and is forwarded to the destination. As the DRE message is received by the distant Cisco WAE, it first parses the message to identify the components that make up the message, including new signatures and data chunks, signatures to redundant data chunks that were removed from transmission, and the message validity signature.

The DRE receiver then updates its DRE database with the new signatures and data chunks contained within the DRE message and then removes the new signatures from the message. Any signatures contained within the DRE message that point to redundant data, or existing entries in the DRE database, are replaced with the actual data chunks that the signatures represent and the signatures are removed from the DRE message. DRE then removes the original message validity signature. At this point, the message should appear as it did before being processed by its peer.

Once the DRE message has been rebuilt, a new message validity signature is computed and compared against the original message validity signature. Should the two message validity signatures be identical, DRE knows that the message has been rebuilt properly using valid data chunks, and can be forwarded to the intended destination. Should the two message validity signatures not match, DRE knows that the message has been rebuilt using data chunks that are not valid for this message stream, and will then fetch the data chunks from its peer.

Figure 5 below shows the advanced compression of Cisco WAAS using DRE and LZ compression:

Figure 5. Cisco WAAS Advanced Compression.



APPLICATION TRAFFIC POLICY ENGINE (ATP)

Cisco WAAS provides administrators with the flexibility they need to define how specific traffic types should be handled by the Cisco WAE. Such definitions include optimizations (DRE, LZ compression, flow optimizations), monitor, and bypass. By default, over 25 application types are identified and over 100 application classifiers are provided, each mapped to a specific set of optimizations to provide the most throughput improvement for that specific application. Figure 6 below shows the application types identified in the default configuration:

Figure 6. Default Cisco WAAS Application Policy.

Common Application Types Optimized By Cisco Wide Area Application Services			
Authentication	Backup	Call Management	Conferencing

Console	Content Management	Directory Services	Enterprise Applications
Enterprise Messaging	File Services	File Transfer	Instant Messaging
Name Services	Network Analysis	Printing	Remote Desktop
Replication Software	Database	Remote Access	Storage Protocols
Streaming	Systems Management	Version Management	Intranet/Internet

The application traffic policy allows administrators to granularly define applications, protocols, and optimizations to apply using the following components:

- Application name – an application name is used to logically group all identifiers that identify traffic for similar application types.
- Traffic classifier – a classifier is used to specify how to identify interesting traffic. Valid traffic classifiers include source or destination IP address or subnet, TCP port, TCP port range, or a specific RPC identifier (UUID).
- Optimization map – a map is used to group a traffic classifier to a specific application and then define the optimizations that should be performed when such traffic is found.

INDUSTRY-LEADING WIDE AREA FILE SERVICES (WAFS)

Being that Cisco WAAS is built using a modular software architecture, many layers of optimizations can be built into a single software platform. Although Cisco WAAS offers the industry's most powerful and flexible application acceleration solution through advanced network optimizations, additional optimizations for specific protocols may be necessary. Cisco WAAS has an extensible software infrastructure that will provide numerous application-specific optimizations in the future, and today provides such optimizations for file services protocols such as the Common Internet File System (CIFS) for Windows and Network File System (NFS) for UNIX environments. Cisco WAAS file services optimizations are based on the industry-leading Wide Area File Services (WAFS) capabilities built from the Cisco WAFS software version 3.0.

Cisco WAAS file services capabilities integrate transparently not only into the packet network, but also into the logical network. No client or server software installation is required to leverage the file services optimizations offered by Cisco WAAS. Cisco WAAS provides the following WAFS capabilities:

- Protocol-specific optimizations – Cisco WAAS examines client-server communications on a message by message basis to fully understand the operations being performed. As such, WAAS can make intelligent decisions on how to most appropriately handle specific messages. In many cases, Cisco WAAS can suppress or respond to messages locally when it is safe to do so. In other cases, certain messages must always traverse the WAN without modification, and WAAS will transfer them using the underlying network optimizations. With protocol-specific optimizations, Cisco WAAS can suppress approximately 70-98% of messages from ever having to utilize the WAN without compromising protocol semantics or correctness.
- Data and metadata caching – Along with protocol-specific optimizations, each application adapter can have its own data and metadata cache. The data and metadata cache are used to hold partially-cached or fully-cached objects, including files, directory information, attributes, and more. By employing a protocol-specific data and metadata

cache, Cisco WAAS can serve usable content to the requesting user upon validation that the content has not been modified since being cached. For scenarios where an object is cached but has been modified, Cisco WAAS can fetch the updated contents using network optimizations such as DRE and LZ compression.

- Centralized file storage – Cisco WAAS allows IT to centralize distributed file servers, storage capacity, and data into the data center where IT staff is readily available. Centralizing distributed servers and storage has many tangible benefits, including:
    - Fewer devices to manage – Cisco WAAS can effectively replace the need for distributed file servers which minimizes the number of devices to manage in the infrastructure. This also eliminates many costly components, including servers, server operating systems and maintenance, OS patching and hotfixes, antivirus, tape drives and libraries, tape cartridges, backup software, and more.
    - Leverage existing data center infrastructure – With a consolidated infrastructure, application and file servers can fully utilize the data center infrastructure components, including server virtualization and storage virtualization.
    - Fewer points of data protection – By consolidating distributed file server storage and data into the data center, fewer copies of data must be protected. This helps control the cost of protecting data and maintaining compliance with federal or industry regulation.
    - Streamlined disaster recovery and business continuity – Cisco WAAS enables consolidation and minimizes the amount of application and file storage infrastructure necessary to support a distributed enterprise. With fewer remote application instances and fewer copies of data, disaster recovery and business continuity planning, deployment, and management are drastically simplified.
- Uncompromised data integrity – Cisco WAAS will only optimize messages and serve data when it is absolutely safe. Critical messages always propagate to the data center without modification and can leverage the network optimizations provided by WAAS. With Cisco WAAS, the data center file server or NAS device always owns the data itself and the state of the data. As such, when a user closes a file and exits the application, the data is safely stored in the data center.
- Integration with Data Redundancy Elimination (DRE) – Cisco WAAS can leverage the advanced compression layers provided by DRE and LZ compression. By integrating with DRE, Cisco WAAS can almost fully decrease the amount of bandwidth consumed by application messaging to the amount of changed data. This is extremely helpful in cases where messages must traverse the WAN or when file data is being written back to the file server unmodified or partially modified.
- Disconnected mode of operation – The WAFS application adapter also provides a read-only disconnected mode of operation for situations where the WAN or the file server or NAS device have gone offline for an extended period of time. For information that needs to be accessible during periods of disconnection, Cisco WAAS will aggressively cache files, folders, metadata information, and access control information. During periods of disconnection, a nearby domain controller can be used to authenticate users and the WAE can validate that requesting users are authorized to access cached data in a read-only fashion.

SUMMARY

Information Technology (IT) faces significant pressure to do more with less: higher application and data availability, higher levels of performance and throughput, fewer people, fewer devices, and less time. Application delivery technologies help IT organizations to consolidate application infrastructure from distributed sites into the data center while providing optimizations necessary to improve application and data access performance over the WAN. Cisco WAAS provides robust optimizations for the network and specifically for applications to help IT improve delivery of applications and application data to the enterprise edge while minimizing infrastructure requirements.

Corporate Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 526-4100

European Headquarters
Cisco Systems International BV
Haarlerbergpark
Haarlerbergweg 13-19
1101 CH Amsterdam
The Netherlands
www-europe.cisco.com
Tel: 31 0 20 357 1000
Fax: 31 0 20 357 1100

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-7660
Fax: 408 527-0883

Asia Pacific Headquarters
Cisco Systems, Inc.
168 Robinson Road
#28-01 Capital Tower
Singapore 068912
www.cisco.com
Tel: +65 6317 7777
Fax: +65 6317 7799

Cisco Systems has more than 200 offices in the following countries and regions. Addresses, phone numbers, and fax numbers are listed on **the Cisco Web site at www.cisco.com/go/offices.**

Argentina • Australia • Austria • Belgium • Brazil • Bulgaria • Canada • Chile • China PRC • Colombia • Costa Rica • Croatia • Cyprus Czech Republic • Denmark • Dubai, UAE • Finland • France • Germany • Greece • Hong Kong SAR • Hungary • India • Indonesia • Ireland Israel • Italy • Japan • Korea • Luxembourg • Malaysia • Mexico • The Netherlands • New Zealand • Norway • Peru • Philippines • Poland Portugal • Puerto Rico • Romania • Russia • Saudi Arabia • Scotland • Singapore • Slovakia • Slovenia • South Africa • Spain • Sweden Switzerland • Taiwan • Thailand • Turkey • Ukraine • United Kingdom • United States • Venezuela • Vietnam • Zimbabwe

Printed in the USA